Risk of lung cancer and physical activity by smoking status and body mass index, the Norwegian Women and Cancer Study

Borch, Kristin Benjaminsen

2019

Borch , K B , Weiderpass , E , Braaten , T , Hansen , M S & Licaj , I 2019 , ' Risk of lung cancer and physical activity by smoking status and body mass index, the Norwegian Women and Cancer Study ' , European Journal of Epidemiology , vol. 34 , no. 5 , pp. 489-498 . https://doi.org/10.1007/s1065

http://hdl.handle.net/10138/301520
https://doi.org/10.1007/s10654-018-0446-0

publishedVersion

European Journal of Epidemiology (2019) 34:489–498
https://doi.org/10.1007/s10654-018-0446-0

CANCER

Risk of lung cancer and physical activity by smoking status and body mass index, the Norwegian Women and Cancer Study

Kristin Benjaminsen Borch[1] · Elisabete Weiderpass[1,2,3,4] · Tonje Braaten[1] · Merethe Selnes Hansen[1] · Idlir Licaj[1,5]

Received: 15 November 2017 / Accepted: 25 September 2018 / Published online: 1 October 2018

Abstract

We aimed to investigate physical activity (PA) and risk of different histological subtypes of lung cancer according to smoking status and body mass index using repeated measurements in a large cohort of women in Norway. The study sample for the multiple imputation analyses consisted of 86,499 and for the complete-case analysis 80,802 women. Repeated measurements of PA level, smoking habits, weight, and height were available for 54,691 women (63.2%), who were included in repeated measurement analyses combined with multiple imputation to address attrition. Cox proportional hazards regression models were used to calculate hazard ratios with 95% confidence intervals. During a median follow-up of 12.9 years, 866 cases of primary lung cancer were identified. We found an inverse association between PA and lung cancer overall. The results were consistent from multiple imputed data analysis to complete-case analysis of PA and possible confounders. We observed a similar trend for adenocarcinoma, but not for squamous cell or small cell carcinomas. Our findings suggest a more pronounced association between lung cancer overall and PA levels in current and former smokers, and in normal-weight and overweight participants with increasing PA levels. The potential of a modifiable lifestyle factor as PA to reduce the risk of lung cancer independently of smoking status is important in public health.

Keywords Lung cancer · Physical activity · Smoking · Prospective study · Women

Introduction

Lung cancer incidence has been increasing among women worldwide and in Norway [1–3]. In Norway, it is the third most common cancer among women, accounting for 10% of all cancers, with 1465 new cases diagnosed in 2016 [3]. Lung cancer is one of the most incurable cancers due to late presentation and disease recurrence with high fatality [4]. Five-year relative survival for lung cancer is low and was 22.0% for Norwegian women in 2012–2016 [3].

Smokers are 14 times more likely to develop lung cancer compared to non-smokers [5]. However, not all smokers develop lung cancer, suggesting individual variability in susceptibility to smoke-related respiratory carcinogens [5]. It is questioned if the increasing incidence in women compared to men is a result of smoking patterns and more susceptibility in developing lung cancer [1, 6]. In a recent study of almost 300,000 Norwegian women findings showed that given the same lifetime exposure women had an increased susceptibility to lung cancer compared to men [7]. Likewise, in a recently published study form the

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10654-018-0446-0) contains supplementary material, which is available to authorized users.

✉ Kristin Benjaminsen Borch
Kristin.benjaminsen.borch@uit.no

[1] Department of Community Medicine, Faculty of Health Sciences, UiT, The Arctic University of Norway, Tromsö, Norway

[2] Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

[3] Department of Research, Cancer Registry of Norway, Institute of Population-Based Cancer Research, Oslo, Norway

[4] Genetic Epidemiology Group, Folkhälsan Research Centre, Faculty of Medicine, University of Helsinki, Helsinki, Finland

[5] Department of Clinical Research, François Baclesse Cancer Centre, Caen, France

United States, they found a higher incidence of lung cancer among young women compared with young men, which are not fully explained by sex differences in smoking behavior [8]. Therefore it is important to investigate whether a modifiable lifestyle factor as physical activity potentially can reduce the risk of lung cancer among women according to smoking habits. Furthermore, there is evidence that increasing body mass index (BMI) is a protective factor against lung cancer [9]. In 2002, the International Agency of Research in Cancer concluded that the association between physical activity (PA) and risk of lung cancer remained inconclusive [10]. In the World Cancer Research Fund/American Institute for Cancer Research (WCRF/AICR) report from 2007 to 2018, the evidence for a protective effect of PA on lung cancer was categorized as limited [11, 12]. However, in the years that followed, several meta-analyses concluded that there was an association between recreational PA and a reduced risk of lung cancer [5, 13–18]. Moreover, it has been suggested that the association between PA and risk of lung cancer among smokers is stronger in women than in men [5]. The WCRF/AICR report included different domains of PA, but few studies have investigated both baseline and repeated measurements of PA and other lifestyle factors [11].

It is still unclear whether the association between PA and lung cancer is the result of an underestimation of lifetime smoking; and therefore a better understanding of this association in never smokers is needed. Indeed, few studies to-date have investigated the association between PA and lung cancer in never smokers, and those that did found no statistically significant association [19–23].

We aimed to investigate PA and risk of different histological subtypes of lung cancer according to smoking status and BMI using repeated measurements of PA and smoking status in a large cohort of women in Norway.

Methods

The Norwegian Women and Cancer Study

The Norwegian Women and Cancer (NOWAC) Study is a nationally representative cohort study that has been described in detail previously [24]. Briefly, random samples of Norwegian women aged 30–70 years were invited to participate during three waves of data collection (1991/1992, 1996/1997, and 2003/2004) [24, 25]. More than 172,000 women were enrolled in the study and completed a questionnaire with detailed questions regarding lifestyle, diet, and health, with an overall response rate of 52.7%. The NOWAC Study was approved by the Regional Committee for Medical Research Ethics and the Norwegian Data Inspectorate, and all participants included in the study gave written informed consent.

In this analysis we used information from enrollment questionnaires completed in 1996–2004 (baseline), and from follow-up questionnaires completed 6–8 years after enrollment (repeated measurement). In total 101,321 women completed baseline questionnaires and were eligible for inclusion in this study. We excluded women with prevalent cancers other than non-melanoma skin cancer at baseline (n = 4450), those who emigrated or died before the start of follow-up (n = 13), and those who had missing information on PA level (n = 9208) and smoking status (n = 1151) at baseline. Thus, the final analytical study sample for chained multiple imputation analyses consisted of 86,499 women. For complete-case baseline and repeated measurement analyses we further excluded women with missing information in pack-years, height, weight, years of education and fruit consumption (n = 5697). The final analytical study sample for complete-case analyses consisted of 80,802 women. Follow-up information on PA level, smoking, weight, and height were available for 54,691 (63.2%) of these women, who were included in the corresponding repeated measurements analyses.

Data collection

The description and validation of the assessment of PA in the NOWAC Study have been described elsewhere [26]. Briefly, respondents reported their PA level in the NOWAC questionnaire on a 10-point scale after reading the following explanation: “By physical activity we mean activity both at work and outside work, at home, as well as training/exercise and other physical activity, such as walking, etc. Please mark the number that best describes your level of physical activity; 1 being very low and 10 being very high”. The scale therefore reflects the amount of PA across different domains, including recreational, occupational, transportation, and household PA, and combines them into one global PA level. This PA scale appeared valid to rank PA level in Norwegian women, but not to quantify a definite dose (i.e. frequency, duration and intensity) of PA [26]. Information on the covariates height, weight, years of education, alcohol consumption, fruit consumption, vegetable consumption, menarche age, menopausal status, number of children, hormone therapy use and oral contraception use were taken from the NOWAC questionnaire. Information on height and weight was used to calculate BMI (kg/m^2). Women also answered questions on smoking status (never, former, or current), duration, age at initiation and number of cigarettes smoked per day. From this information number of pack-years was calculated as number of cigarettes smoked per day divided by 20 and multiplied by years of smoking at baseline and at

follow-up. Women who reported that they were current or former smokers at baseline and never smoker at follow-up were categorized as former smokers at follow-up.

Women diagnosed with a primary, invasive, malignant neoplasm of the lung (International Statistical Classification of Diseases, Injuries and Causes of Death Revision 10 codes C33-34) [27] were identified through linkage to the Cancer Registry of Norway, from which date of diagnosis and morphology (International Classification of Diseases for Oncology, 3rd edition) were also obtained. Based on morphology, lung cancers were categorized into adenocarcinoma, squamous cell carcinoma, small cell carcinoma, large cell carcinoma, other non-small cell carcinoma, and other or not otherwise specified carcinoma. Here we present data on the risk of lung cancer overall and on adenocarcinoma, squamous cell carcinoma, and small cell carcinoma, as they were the most frequent subtypes of lung cancer. Information on date of death or emigration was obtained through linkage to the Norwegian National Population Register.

Statistical methods

All the analyses and multiple imputations were done in STATA version 14.0 (Stata Corp, College Station, TX, USA).

We used the method proposed by Hu et al. [28], i.e. baseline data was used until follow-up information became available, death, or emigration, whichever occurred first. Thereafter, follow-up information was applied until death, emigration, or the end of the study period, whichever occurred first. In the analysis using repeated PA measurements, we also used follow-up information on BMI, smoking status, and fruit consumption once it became available. We used Cox proportional hazards regression models to calculate hazard ratios (HRs) with 95% confidence intervals (CIs) comparing five categories of PA level (1–2, 3–4, 5–6, 7–8, and 9–10). PA level 5–6 was set as the reference group. Follow-up time was defined as the interval between age at baseline and age at cancer diagnosis, death, emigration, or the end of follow-up (31 December 2014), whichever came first.

Departures from the proportional hazards assumption in the Cox models were tested through the inclusion of an interaction variable between categories of PA level and underlying time (age) [29]. A preliminary analysis of baseline data was used to select the covariates [height (meters), weight (kg), years of education, smoking status (never, former, current), number of pack-years, alcohol consumption (g/day), fruit consumption (g/day), vegetable consumption (g/day), menarche age (years), menopausal status (pre, peri, postmenopausal), number of children, hormone therapy use (yes/no) and oral contraception use (yes/no)] for which we adjusted in the models at complete-case analyses and imputed analyses. The covariates that led to a change of at least 10% in the PA regression coefficient were included in the final models: BMI (normal weight: < 25, overweight: 25–29.9, obese: ≥ 30 kg/m^2), years of education (≤ 9, 9–12, ≥ 13 years), smoking status in combination with pack-years (never, former and pack-years < 10, former and pack-years ≥ 10, current and pack-years < 10 and current and pack-years ≥ 10, and fruit consumption (g/day) in quartiles. In all the analysis models we stratified by birth cohort (stratum 1: enrolled in 1996–1997 and born 1927–1942; stratum 2: enrolled in 1996–1997 and born 1943–1957; and stratum 3: enrolled in 2003–2007 and born 1943–1957). To test for linear trend, we used the original, 10-point PA scale, modelled as a continuous variable in the analyses. Interactions between PA and the above-mentioned categories of BMI, years of education, and smoking status were tested using likelihood ratio statistics.

In order to counteract residual confounding due to smoking and BMI, we repeated our analyses in the stratified samples by these two factors and with the same adjustments as in the main models. In the stratified analyses, PA levels 7–10 were collapsed both in baseline and repeated measurements analyses due to the low number of participants in the separate groups. In never smokers we did not adjust for pack-years. We also conducted complete-case baseline sensitivity analysis in which we considered only information at cohort entrance [28].

Multiple imputation

Under the assumptions that data was missing at random, we performed chained multiple imputation to deal with the missing information at baseline and follow-up [30]. The pattern of missing was confirmed arbitrary. We used a similar imputation process as described in previous publications [31–33]. To reduce sampling variability, we created 20 replicate datasets from the imputation simulation [34]. We used separate imputation models for the outcomes lung cancer overall and for each histological subtypes adenocarcinoma, squamous cell carcinoma, and small cell carcinoma, including all of the variables from the final analysis; PA, smoking status and number of pack-years, BMI and fruit consumption (baseline and follow-up information) and education (baseline information). We further included additional baseline variables in the imputation process: age at smoking initiation, alcohol consumption, number of children and menopausal status. These variables were only used in the imputation process in order to increase the predictive power of the imputation procedure. The Nelson–Aalen cumulative hazard estimator was included as a predictor in the imputation models [30].

The estimates from the 20 imputed datasets were combined using Rubin's rules [35].

Results

During more than 1.1 million person-years and a median follow-up of 12.9 years, 866 cases of primary lung cancer overall were identified. At baseline there were 353 (44.1%) adenocarcinomas, 125 (14.4%) squamous cell carcinomas, and 165 (18.1%) small cell carcinomas (Table 1). Other histological subtypes represented 22% of the cases. Mean age at cohort entrance was 51.7 years and the mean age at lung cancer diagnosis was 64.4 years. At baseline, 43% of the women reported a PA level of 5–6, and in total nearly 74.0% reported a PA level of 5 or higher. Sixty-one participants (7%) later identified with lung cancer were never smokers at baseline, whereas 665 (76.8%) were current smokers. Women with a PA level of 5 or higher were more frequently never smokers, had a lower BMI, and a higher fruit and vegetable consumption than their counterparts with lower PA levels (Table 1). Among 20,622 never smoking women 1279 (6.25%) and 42 (0.34%) were former and current smokers at follow-up, respectively. Among current smokers at baseline 3999 (19.5%) had stopped smoking at follow-up. Women who dropped out of the study were more often current smokers, but on average they had PA levels that were similar to those of women

Table 1 Selected participant characteristics at study enrollment (baseline) in the Norwegian Women and Cancer study by physical activity (PA) level (N = 86,499)

Characteristic	PA level					
	1–2 N = 4079 (4.7%)	3–4 N = 19,005 (22%)	5–6 N = 37,154 (43%)	7–8 N = 21,819 (25.2%)	9–10 N = 4442 (5.1%)	Total N = 86,499 (100%)
Age (mean, ± SE)	53.2 (0.1)	52.2 (0.05)	51.5 (0.03)	51.3 (0.04)	51.9 (0.10)	51.7 (0.02)
Follow-up time years (mean, ± SE)	12.8 (0.07)	13.1 (0.03)	13.0 (0.02)	12.7 (0.03)	12.7 (0.06)	12.9 (0.01)
Lung cancer overall (n, %)	83	234	354	149	46	866 (100%)
Adenocarcinoma[a] (n, %)	26	114	154	67	21	382 (44.1%)
Squamous cell carcinoma[a] (n, %)	11	29	58	19	8	125 (14.4%)
Small cell carcinoma[a] (n, %)	22	37	65	32	9	165 (19.1%)
Other histological subtypes[b] (n, %)	24	54	77	31	8	194 (22%)
BMI (mean, ± SE)[c]	26.9 (0.09)	25.8 (0.03)	24.6 (0.02)	23.8 (0.02)	23.6 (0.05)	24.7 (0.01)
Years of education (mean, ± SE)[d]	11.4 (0.06)	12.1 (0.03)	12.3 (0.02)	12.6 (0.02)	11.8 (0.06)	12.3 (0.01)
Smoking status (%)[e]						
Never	29.9	36.6	37.7	37.8	35.8	37
Former	30.2	31.7	33.2	35.5	33.5	33.3
< 10 pack-years	66.1	74	77	77.8	79.8	76.3
≥ 10 pack-years	33.9	26	23	22.2	20.2	23.7
Current	39.9	31.8	29.2	26.7	30.7	29.7
< 10 pack-years	22.9	23.7	27.8	26.6	32.2	27.2
≥ 10 pack-years	77.1	76.3	72.2	70.4	67.8	72.8
Pack-years of smoking in ever smokers (mean, ± SE)	13.6 (0.20)	11.3 (0.09)	10.0 (0.06)	9.4 (0.07)	9.5 (0.16)	10.3 (0.04)
Alcohol consumption, mean ± SE (g/day)[f]	3.6 (0.11)	3.6 (0.04)	3.5 (0.03)	3.7 (0.04)	3.4 (0.10)	3.6 (0.02)
Fruit consumption, mean ± SE (g/day)	166 (2.35)	185 (1.04)	204 (0.76)	225.2 (1.07)	240.4 (2.70)	205.2 (0.52)
Vegetable consumption, mean ± SE (g/day)	120.5 (1.4)	131.2 (0.62)	140.1 (0.45)	154.5 (0.65)	166.1 (1.7)	142.2 (0.31)

SE standard error, *BMI* body mass index

[a]Histological subtypes of lung cancer: only the main subtypes are included in the analysis

[b]Other histological subtypes: large cell, other non-small cell and other are not included in the analysis

[c]Number of total missing in BMI 1567 (1.8%)

[d]Number of total missing in years of education 4267 (4.9%)

[e]Number of total missing in smoking status 1151 (1.4%)

[f]Number of total missing in alcohol consumption 2095 (2.6%)

who did not drop out of the study. There was four missing observations in pack-years at baseline among former and current smokers. The pattern of missing from baseline to follow-up shows that 63.2% (54,691) of the participants had information at follow-up for smoking status, weight, and height. The proportion of missing at follow-up reported in Table 2 shows that there was 41.8% missing in PA, 37.9% in BMI, 37.2% in smoking status and 42.1% in pack-years (former and current smokers). However, the missing is arbitrary and equally distributed as the complete data over levels of PA (Table 2), by smoking status and BMI groups (Supplementary Tables 3 and 4).

In the analysis using the imputed dataset and adjusted models we observed that low PA level was negatively associated with the risk of lung cancer overall (*P for trend* = 0.000) (Table 3). Participants with a PA level of 1–2 had 73% increased risk of lung cancer overall compared to those with PA levels of 5–6 (HR = 1.73; 95% CI 1.30–2.30). Participants with a PA level higher than 5–6 showed a tendency for a reduced risk of lung cancer overall (HR for PA level 7–8 = 0.84; 95% CI 0.67–1.06 and HR for PA level 9–10 = 0.87; 95% CI 0.57–1.32), however not significant. The results for adenocarcinoma showed a comparable statistically significant trend; but no significant associations were found for any of the other investigated histological subtypes (Table 3).

Interactions between PA and categories of BMI, smoking status, and years of education were not significant (*P for trend* = 0.87, *P for trend* = 0.42, and *P for trend* = 0.24, respectively). In spite of this, we investigated the association between PA and risk of lung cancer overall in analyses stratified by smoking status and BMI groups. Analyses stratified by smoking status showed no significant association between levels of PA and lung cancer overall in never smokers. However, in former and current smokers with PA levels 1–2 there was a significant, increased risk of lung cancer overall (HR = 2.34, 95% CI 1.42–3.85 and HR = 1.74, 95% CI 1.29–2.33, respectively), with a consistent decreasing risk with increasing PA levels (Table 3). Analyses stratified by BMI showed a higher risk of lung cancer overall among normal-weight (HR = 1.60, 95% CI 1.12–2.28) for PA levels 1–2, and more pronounced for overweight participants with PA levels 1–2 (HR = 2.37, 95% CI 1.54–3.63) (Table 3). No significant associations were found for obese participants.

Complete-case analyses yielded similar results as the imputed data analysis (Supplementary Table 2) in the main analysis. However, the results for the analysis stratified by smoking status was not statistically significant in former smokers. Stratification for BMI yielded results similar to those of imputed data analyses for normal-weight. For overweight participants the association between PA levels and lung cancer overall was not significant in the complete case analysis (Supplementary Table 2).

Discussion

In this large Norwegian cohort, we found an inverse association between self-reported PA level and risk of lung cancer overall using chained multiple imputations. The results were consistent when using complete-case analysis baseline and repeated measurements of PA, BMI and smoking status. We observed a similar trend for adenocarcinomas; however no statistically significant associations were found for squamous cell carcinoma or small cell carcinoma. Our findings further suggest a more pronounced association between lung cancer overall and PA levels in former and current smokers, with an increased risk corresponding to low levels of PA, whereas no significant associations were found in never smokers. We found a

Table 2 Distribution of missing information (%) by physical activity level at baseline, the Norwegian Women and Cancer Study, 1996–2014

Characteristics	Physical activity level at baseline (%)					
	1–2 N = 4079 (4.7%)	3–4 N = 19,005 (22%)	5–6 N = 37,154 (43%)	7–8 N = 21,819 (25%)	9–10 N = 4442 (5.1%)	Total N = 86,499 (100%)
Physical activity level at follow-up	47.3	41.1	41.1	41.9	45.7	36,194 (41.8)
Body mass index at baseline	2.1	2.0	1.8	1.6	2.1	1567 (1.8)
Body mass index at follow-up	42.0	37.3	37.5	38.2	40.0	32,822 (37.9)
Smoking status at follow-up	41.0	36.4	36.8	37.5	39.5	32,191 (37.2)
Number of pack-years at follow-up[a]	46.6	41.7	41.9	43.7	45.7	23,290 (42.1)
Duration of education (years)	6.6	5.2	4.6	4.9	6.0	4267 (4.9)

Missing in pack-year at baseline was 4 observations and is not included in the table

[a]Number of pack-year only in former and current smokers at follow-up (n = 86,499)

Table 3 Multivariable hazard ratios and 95% confidence intervals in *chained multiple imputation* analyses of the association between physical activity (PA) and lung cancer including baseline and repeated measurements in the Norwegian Women and Cancer Study (n = 86,499)

	Number of lung cancer cases	PA levels					
		1–2	3–4	5–6	7–8	9–10	*P* trend
Lung cancer overall[a]	866	1.73 (1.30–2.30)	1.33 (1.08–1.64)	1.00	0.84 (0.67–1.06)	0.87 (0.57–1.32)	0.000
Adenocarcinoma[a]	382	1.50 (0.94–2.39)	1.42 (1.06–1.90)	1.00	0.85 (0.61–1.19)	0.60 (0.27–1.31)	0.001
Squamous carcinomas[a]	125	1.48 (0.71–3.11)	1.29 (0.78–2.14)	1.00	0.88 (0.47–1.66)	1.06 (0.38–2.95)	0.163
Small cell[a]	165	1.82 (1.01–3.29)	0.93 (0.57–1.52)	1.00	0.80 (0.48–1.35)	1.16 (0.53–2.52)	0.156
Lung cancer overall stratified analyses by							
Smoking	Never n = 61	1.39 (0.41–4.71)	1.23 (0.63–2.42)	1.00	1.26 (0.66–2.40)	0.45 (0.06–3.34)	0.600
	Former n = 140	2.34 (1.42–3.85)	1.63 (1.16–2.31)	1.00	0.96 (0.65–1.41)	1.08 (0.55–2.11)	0.000
	Current n = 665	1.74 (1.29–2.33)	1.29 (1.04–1.59)	1.00	0.79 (0.61–1.01)	0.89 (0.57–1.38)	0.000
		1–2	3–4	5–6	7–10		*P* trend
Body mass index	Normal n = 608	1.60 (1.12–2.28)	1.39 (1.12–1.73)	1.00	0.84 (0.67–1.05)		0.000
	Overweight n = 196	2.37 (1.54–3.63)	1.32 (0.95–1.83)	1.00	1.03 (0.71–1.50)		0.001
	Obese n = 62	1.94 (0.98–3.84)	1.38 (0.75–2.52)	1.00	0.66 (0.26–1.69)		0.012

Multiple imputation of covariates in the multivariable analyses conducted with chained equation. 20 imputed data sets were generated

[a]Model 1: Multivariable model adjusted for body mass index, years of education, smoking status and pack-years (never; former and < 10 pack-years; former and ≥ 10 pack-years; current and < 10 pack-years; current and ≥ 10 pack-years), fruit consumption (g/day). Model stratification by birth cohort (strata 1: enrolled in 1996–1997 and born 1927–1942; strata 2: enrolled in 1996–1997 and born 1943–1957; and strata 3: enrolled in 2003–2007 and born 1943–1957)

reduced risk of lung cancer in normal-weight and overweight participants with increasing PA.

The heterogeneity between studies in the relationship between lung cancer and PA level is challenging, as the type of PA measured, the lack of dose–response data, study design, sex, and sample size make it difficult to compare these studies and to calculate overall estimates. Most studies that investigated the association between PA and the risk of lung cancer in women focused on recreational PA [16, 17, 36]; fewer studies have captured total PA (including recreational, occupational, household, and transportation) [19, 22, 37–39]. In its Second Expert Report from 2007 to 2018, the WCRF/AICR concluded that the evidence for an association between PA and lung cancer is "limited-suggestive" [11, 12]. The five cohort studies included in that comprehensive report all had overall summaries of PA (recreational and non-recreational), but the lack of detail in the evidence made dose–response analyses impossible, and only two of the studies included women [11]. The NIH-AARP Diet and Health Study by Leitzmann and colleagues found an inverse association between total PA (both recreational and non-recreational) and risk of lung cancer in analyses adjusted for sex, and these findings were consistent in sex-specific analyses and over other covariates, i.e. age, education level, and BMI [22]. Our results showed a significant decreased risk of lung cancer overall, and are thus similar to those of the NIH-AARP Diet and Health Study; however the measurements of PA between the studies are not comparable [22]. The results of the two most recent meta-analyses on recreational PA and risk of lung cancer showed statistically significant, inverse associations in the range of 13–25% for lung cancer from cohort studies [16, 17]; however these estimates included several studies that enrolled only men. One meta-analysis that estimated the association among women separately found an overall relative risk of 0.73 (95% CI 0.63–0.86); however it included both case–control and cohort studies, and the heterogeneity between the studies was found to be high (I-squared = 50.9%) [16]. Based on cohort studies that included sex-adjusted analysis and/or specific analysis of women, five found inverse associations [21, 22, 40–42], while 12 others [19, 23, 36–39, 43–49] found no association between the risk of lung cancer and PA level (including both recreational and non-recreational PA). The suggested mechanisms of PA in protecting against lung cancer is through increase in enzymatic systems that detoxify chemical carcinogens and in this way protect the lungs [50, 51]. In addition, increase in antioxidants levels in PA have been suggested to explain the protective effect of PA [49]. In a nested case–control study within EPIC, investigating endogenous antioxidant biomarkers in the association

between recreational PA and lung cancer suggesting that PA is protective against lung cancer risk in former- and non-smokers [49].

The investigation of the association between PA and lung cancer is complicated by tobacco smoking, which acts as a powerful confounder in the causation of lung cancer. The protective association of PA observed among smokers may be confounded by the amount and duration of cigarette smoking [52]. The solution to this problem may be to investigate this association among never smokers only [23]. However, that could represent a selected group bias, with participants who have a low prevalence of other exposures of relevance [53]. A recent meta-analyses stratified by smoking behavior to address this potential bias, and concluded that PA was not related to lung cancer in never smokers and that there was a reduced risk in former and current smokers [52]. Our stratified analyses indicated that the risk of lung cancer overall was reduced among current and former smokers with increasing PA levels. We found no significant association between lung cancer overall and PA level in never smokers. These results are consistent with findings from several other studies [21–23, 38, 52, 54]. Today, there are no predominant factor that can fully explain lung cancer in never smokers, but risk factors considered important include second hand smoking, radon exposure, environmental exposures, history of lung diseases and genetic factors [55]. Further, it has been suggested that the etiology of lung cancer among never smokers may be distinct from that of smokers [55] which justifies to do separate analysis by never-smokers and ever-smokers. Therefore, the finding of no association in the absence of cigarette smoking can be related to residual confounding. Furthermore, the relatively low number of lung cancer cases among never smokers made these analyses less robust, and must therefore be interpreted with caution. Although we found an association between PA and smoking in the stratified analyses the residual confounding of smoking due to imprecisely measured smoking behavior cannot be ruled out.

The reduced risk of lung cancer was more profound among normal-weight and overweight women with increasing PA levels in our study. This corresponds with findings from other studies [22, 40, 41, 54], which showed that low and medium BMI in those with higher PA levels reduced the risk of lung cancer. Other cohort studies adjusted for BMI and did not report group-specific analyses [19, 21, 23, 37–39, 48].

The strengths of our study include its prospective, population-based design and the use of a high-quality, national cancer registry to identify lung cancer cases [24]. The presence of repeated measurements on the exposure and potential confounders is also a considerable strength. This justify the importance of using repeated measurements in our study where almost 20% of the women changed smoking status from current to former smokers during follow-up, despite the amount of missing which were addressed with advanced chained multiple imputation methods. Rubin's rule were applied to take into account the variability in the results using 20 imputed datasets, to reflect the uncertainty that are associated with missing values. The prospective design precluded bias attributable to recall bias of PA by participants independently of lung cancer. Moreover, the PA scale we used has been validated [26] and correlated well with all-cause mortality rates [56]. Our assessment of PA in the NOWAC Study comprised total PA, covering the domains of recreation, occupation, household and transportation, with one repeated measurement during follow-up. We chose to impute missing information at baseline and follow-up, assuming a missing-at-random framework. Our chained imputation models included all variables for the final Cox proportional hazards regression models and additional variables in order to predict incomplete variables or to predict whether the incomplete variable was missing [57]. This method has been developed and proved robust in several of studies within the NOWAC study [31–33, 58], and the consistency with the complete-case analysis strengthens the results. Nevertheless, we cannot rule out that some of the information is still missing-not-at-random which may lead to biased estimations. Our large prospective study included a high number of lung cancer cases, and thus it was possible to investigate histological subtypes. In Norway the mean age at smoking initiation has been found to be $\leq$ 20 years [7, 59]. The 30-year lag period between smoking initiation and time of cohort enrollment (mean age 52 years) for the majority of smokers gives an increased risk of lung cancer. The follow-up time (13 years) in our study adds to this, justifying a relatively long follow-up.

The total self-reported measure of PA in the NOWAC Study cannot differentiate intensity, duration, and frequency of PA, nor the type of PA, and given the self-reported nature of this variable, measurement errors cannot be ruled out. A known problem in self-reporting of a desirable behaviour like PA is the tendency to overestimate; i.e. report levels that are higher than one really has [60, 61]. However, measurement errors would likely lead to a non-differential bias in relation to the cancer versus non-cancer status and a potential underestimation of the true effect. The PA assessment used in this study may not apply to women in other countries, and the scale cannot directly be compared to other studies using a different measure of PA. Moreover, the potential for residual confounding, in particular by lifetime smoking habits (i.e. age at initiation, time since quitting for former smokers, type of cigarettes smoked and passive smoking), is a recurrent problem in studies of lung cancer and is possible affecting

our results. In particular, the PA measurement could introduce information bias, together with other risk factors that were not measured; i.e. second-hand smoking and radon exposure. Certain other lifestyle-related factors may act as confounders or effect-modifiers and cannot be excluded.

Conclusions

In this study we used a large nationally representative and prospective cohort and presented analyses using time varying covariates and dealt with attrition using chained multiple imputations. We found that higher PA was associated with lower risk of lung cancer overall and no association among never smokers. Women who were within the normal- and overweight range appeared to benefit from a higher protective effect of PA, as did those who were current and former smokers. However, the findings must be interpreted with caution as residual confounding by smoking behavior cannot be ruled out. The continuing increase in lung cancer among women in Norway, as well as in a numerous of countries may foreshadow a higher future burden of lung cancer in women than previously expected. Thus, there is a need to intensify anti-tobacco measures to decrease smoking in women. To promote smoking cessation and prevent people to start smoking is the primary means of preventing lung cancer. Furthermore, to prevent lung cancer in all groups of the population independently of their smoking status through a modifiable lifestyle factor as PA is of tremendous interest in the field of public health.

Funding This project received financial support from the Norwegian Extra Foundation for Health and Rehabilitation through EXTRA funds, Grant No. 2012/2/0048-1220048001.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

References

1. Graham PD, Thigpen SC, Geraci SA. Lung cancer in women. South Med J. 2013;106(10):582–7.
2. Torre LA, Siegel RL, Ward EM, Jemal A. Global cancer incidence and mortality rates and trends—an update. Cancer Epidemiol Biomark Prev. 2016;25(1):16–27. https://doi.org/10.1158/1055-9965.Epi-15-0578.
3. Cancer Registry of Norway. Cancer in Norway 2015—cancer incidence, mortality, survival and prevalence in Norway. Oslo: Cancer Registry of Norway; 2016.
4. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer. 2015;136(5):E359–86. https://doi.org/10.1002/ijc.29210.
5. Buffart LM, Singh AS, van Loon ECP, Vermeulen HI, Brug J, Chinapaw MJM. Physical activity and the risk of developing lung cancer among smokers: a meta-analysis. J Sci Med Sport. 2014;17(1):67–71. https://doi.org/10.1016/j.jsams.2013.02.015.
6. Khuder SA, Mutgi AB. Effect of smoking cessation on major histologic types of lung cancer. Chest. 2001;120(5):1577–83. https://doi.org/10.1378/chest.120.5.1577.
7. Hansen MS, Licaj I, Braaten T, Langhammer A, Le Marchand L, Gram IT. Sex differences in risk of smoking—associated lung cancer: results from a Cohort of 600,000 Norwegians. Am J Epidemiol. 2017;187(5):971–81. https://doi.org/10.1093/aje/kwx339.
8. Jemal A, Miller KD, Ma J, Siegel RL, Fedewa SA, Islami F, et al. Higher lung cancer incidence in young women than young men in the United States. N Engl J Med. 2018;378(21):1999–2009. https://doi.org/10.1056/NEJMoa1715907.
9. Duan P, Hu C, Quan C, Yi X, Zhou W, Yuan M et al. Body mass index and risk of lung cancer: systematic review and dose-response meta-analysis. Sci Rep. 2015;5:16938. https://doi.org/10.1038/srep16938. https://www.nature.com/articles/srep16938#supplementary-information.
10. Vainio H, Bianchini F. IARC Handbooks of cancer prevention. Volume 6: weight control and physical activity. IARC Handbooks of Cancer Prevention. France: IARC Press; 2002.
11. World Cancer Research Fund/American Institute for Cancer Research. Food, nutrition, physical activity, and the prevention of cancer: a global perspective. Washington DC: AICR; 2007.
12. World Cancer Research Fund/American Institute for Cancer Research. Diet, Nutrition, Physical Activity, and Cancer: a Global perspective. Continuous Update Project Expert Report; 2018. http://www.dietandcancerreport.org.
13. Tardon A, Lee WJ, Delgado-Rodriguez M, Dosemeci M, Albanes D, Hoover R, et al. Leisure-time physical activity and lung cancer: a meta-analysis. Cancer Causes Control. 2005;16(4):389–97. https://doi.org/10.1007/s10552-004-5026-9.
14. Koutsokera A, Kiagia M, Saif MW, Souliotis K, Syrigos KN. Nutrition habits, physical activity, and lung cancer: an authoritative review. Clin Lung Cancer. 2013;14(4):342–50. https://doi.org/10.1016/j.cllc.2012.12.002.
15. Sun J-Y. Physical activity and risk of lung cancer: a meta-analysis of prospective cohort studies. Asian Pac J Cancer Prev. 2012;13(7):3143–7.
16. Zhong S, Ma T, Chen L, Chen W, Lv M, Zhang X, et al. Physical activity and risk of lung cancer: a meta-analysis. Clin J Sport Med. 2016;26(3):173–81.
17. Brenner DR, Yannitsos DH, Farris MS, Johansson M, Friedenreich CM. Leisure-time physical activity and lung cancer risk: a systematic review and meta-analysis. Lung Cancer. 2016;95:17–27. https://doi.org/10.1016/j.lungcan.2016.01.021.
18. Emaus A, Thune I. Physical activity and lung cancer prevention. In: Courneya KS, Friedenreich CM, editors. Physical activity and cancer. Berlin: Springer; 2011. p. 101–33.
19. Sprague BL, Trentham-Dietz A, Klein BEK, Klein R, Cruickshanks KJ, Lee KE, et al. Physical activity, white blood cell count, and lung cancer risk in a prospective cohort study. Cancer

Epidemiol Biomark Prev. 2008;17(10):2714–22. https://doi.org/10.1158/1055-9965.epi-08-0042.
20. Kubík AK, Zatloukal P, Tomášek L, Pauk N, Havel L, Krepela E, et al. Dietary habits and lung cancer risk among non-smoking women. Eur J Cancer Prev. 2004;13(6):471–80.
21. Sinner P, Folsom AR, Harnack L, Eberly LE, Schmitz KH. The association of physical activity with lung cancer incidence in a cohort of older women: the Iowa Women's Health Study. Cancer Epidemiol Biomark Prev. 2006;15(12):2359–63. https://doi.org/10.1158/1055-9965.epi-06-0251.
22. Leitzmann MF, Koebnick C, Abnet CC, Freedman ND, Park Y, Hollenbeck A, et al. Prospective study of physical activity and lung cancer by histologic type in current, former, and never smokers. Am J Epidemiol. 2009;169(5):542–53. https://doi.org/10.1093/aje/kwn371.
23. Lam TK, Moore SC, Brinton LA, Smith L, Hollenbeck AR, Gierach GL, et al. Anthropometric measures and physical activity and the risk of lung cancer in never-smokers: a prospective cohort study. PLoS ONE. 2013;8(8):e70672. https://doi.org/10.1371/journal.pone.0070672.
24. Lund E, Dumeaux V, Braaten T, Hjartåker A, Engeset D, Skeie G, et al. Cohort profile: The Norwegian Women and Cancer Study (NOWAC) Kvinner og kreft. Int J Cancer. 2008;37(1):36–41. https://doi.org/10.1093/ije/dym137.
25. Lund E, Kumle M, Braaten T, Hjartåker A, Bakken K, Eggen A, et al. External validity in a population-based national prospective study—the Norwegian Women and Cancer Study (NOWAC). Cancer Causes Control. 2003;14(10):1001–8. https://doi.org/10.1023/B:CACO.0000007982.18311.2e.
26. Borch KB, Ekelund U, Brage S, Lund E. Criterion validity of a 10-category scale for ranking physical activity in Norwegian women. Int J Behav Nutr Phys Act. 2012;9(1):2.
27. WHO. International statistical classification of diseases and related health problems 10th Revision. http://apps.who.int/classifications/icd10/browse/2015/en. 2010. Accessed 13 Nov 2017.
28. Hu P, Tsiatis AA, Davidian M. Estimating the parameters in the Cox model when covariate variables are measured with error. Biometrics. 1998;54(4):1407–19. https://doi.org/10.2307/2533667.
29. Thiébaut ACM, Bénichou J. Choice of time-scale in Cox's model analysis of epidemiologic cohort data: a simulation study. Stat Med. 2004;23(24):3803–20. https://doi.org/10.1002/sim.2098.
30. White IR, Royston P. Imputing missing covariate values for the Cox model. Stat Med. 2009;28(15):1982–98. https://doi.org/10.1002/sim.3618.
31. Borch KB, Weiderpass E, Braaten T, Jareid M, Gavrilyuk OA, Licaj I. Physical activity and risk of endometrial cancer in the Norwegian Women and Cancer (NOWAC) study. Int J Cancer. 2017;140(8):1809–18. https://doi.org/10.1002/ijc.30610.
32. Licaj I, Sandin S, Skeie G, Adami H-O, Roswall N, Weiderpass E. Alcohol consumption over time and mortality in the Swedish Women's Lifestyle and Health cohort. BMJ Open. 2016. https://doi.org/10.1136/bmjopen-2016-012862.
33. Lukic M, Licaj I, Lund E, Skeie G, Weiderpass E, Braaten T. Coffee consumption and the risk of cancer in the Norwegian Women and Cancer (NOWAC) Study. Eur J Epidemiol. 2016;31(9):905–16. https://doi.org/10.1007/s10654-016-0142-x.
34. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ. 2009;338:b2393. https://doi.org/10.1136/bmj.b2393.
35. Rubin DB. Multiple imputation after 18 + years. J Am Stat Assoc. 1996;91(434):473–89. https://doi.org/10.2307/2291635.
36. Patel AV, Carter BD, Stevens VL, Gaudet MM, Campbell PT, Gapstur SM. The relationship between physical activity, obesity, and lung cancer risk by smoking status in a large prospective cohort of US adults. Cancer Causes Control. 2017;28(12):1357–68. https://doi.org/10.1007/s10552-017-0949-0.
37. Inoue M, Yamamoto S, Kurahashi N, Iwasaki M, Sasazuki S, Tsugane S. Daily total physical activity level and total cancer risk in men and women: results from a large-scale population-based cohort study in Japan. Am J Epidemiol. 2008;168(4):391–403. https://doi.org/10.1093/aje/kwn146.
38. Thune I, Lund E. The influence of physical activity on lung-cancer risk: a prospective study of 81,516 men and women. Int J Cancer. 1997;70:57–62. https://doi.org/10.1002/(SICI)1097-0215(19970106)70:1%3c57:AID-IJC9%3e3.0.CO;2-5.
39. Steindorf K, Friedenreich C, Linseisen J, Rohrmann S, Rundle A, Veglia F, et al. Physical activity and lung cancer risk in the European Prospective Investigation into Cancer and Nutrition Cohort. Int J Cancer. 2006;119(10):2389–97. https://doi.org/10.1002/ijc.22125.
40. Wang A, Qin F, Hedlin H, Desai M, Chlebowski R, Gomez S, et al. Physical activity and sedentary behavior in relation to lung cancer incidence and mortality in older women: the Women's Health Initiative. Int J Cancer. 2016;139(10):2178–92. https://doi.org/10.1002/ijc.30281.
41. Moore SC, Lee I, Weiderpass E, et al. Association of leisure-time physical activity with risk of 26 types of cancer in 1.44 million adults. JAMA Intern Med. 2016. https://doi.org/10.1001/jamainternmed.2016.1548.
42. Hållmarker U, James S, Michaëlsson K, Ärnlöv J, Sandin F, Holmberg L. Cancer incidence in participants in a long-distance ski race (Vasaloppet, Sweden) compared to the background population. Eur J Cancer. 2015;51(4):558–68. https://doi.org/10.1016/j.ejca.2014.12.009.
43. Pukkala E, Poskiparta M, Apter D, Vihko V. Life-long physical activity and cancer risk among Finnish female teachers. Eur J Cancer Prev. 1993;2(5):369–76.
44. Steenland K, Nowlin S, Palu S. Cancer incidence in the National Health and Nutrition Survey I. Follow-up data: diabetes, cholesterol, pulse and physical activity. Cancer Epidemiol Biomark Prev. 1995;4(8):807–11.
45. Alfano CM, Klesges RC, Murray DM, Bowen DJ, McTiernan A, Vander Weg MW, et al. Physical activity in relation to all-site and lung cancer incidence and mortality in current and former smokers. Cancer Epidemiol Biomark Prev. 2004;13(12):2233–41.
46. Bak H, Christensen J, Thomsen BL, Tjønneland A, Overvad K, Loft S, et al. Physical activity and risk for lung cancer in a Danish cohort. Int J Cancer. 2005;116(3):439–44. https://doi.org/10.1002/ijc.21085.
47. Schnohr P, Grønbæk M, Petersen L, Hein HO, Sørensen TI. Physical activity in leisure-time and risk of cancer: 14-year follow-up of 28,000 Danish men and women. Scand J Public Health. 2005;33(4):244–9. https://doi.org/10.1080/14034940510005752.
48. Land SR, Liu Q, Wickerham DL, Costantino JP, Ganz PA. Cigarette smoking, physical activity, and alcohol consumption as predictors of cancer incidence among women at high risk of breast cancer in the NSABP P-1 trial. Cancer Epidemiol Biomark Prev. 2014;23(5):823–32. https://doi.org/10.1158/1055-9965.epi-13-1105-t.
49. Rundle A, Richie J, Steindorf K, Peluso M, Overvad K, Raaschou-Nielsen O, et al. Physical activity and lung cancer among non-smokers: a pilot molecular epidemiological study within EPIC. Biomarkers. 2010;15(1):20–30. https://doi.org/10.3109/13547500903186452.
50. Duncan K, Harris S, Murray Ardies C. Running exercise may reduce risk for lung and liver cancer by inducing activity of antioxidant and phase II enzymes. Cancer Lett. 1997;116(2):151–8. https://doi.org/10.1016/S0304-3835(97)00189-4.

51. Rundle AG, Orjuela M, Mooney L, Tang D, Kim M, Calcagnotto A, et al. Preliminary studies on the effect of moderate physical activity on blood levels of glutathione. Biomarkers. 2005;10(5):390–400. https://doi.org/10.1080/13547500500272663.
52. Schmid D, Ricci C, Behrens G, Leitzmann MF. Does smoking influence the physical activity and lung cancer relation? A systematic review and meta-analysis. Eur J Epidemiol. 2016;31(12):1173–90. https://doi.org/10.1007/s10654-016-0186-y.
53. Adami H-O, Hunter DJ, Trichopoulos D. Textbook of cancer epidemiology. 2nd ed. New York: Oxford University Press; 2008.
54. Mao Y, Pan S, Wen SW, Johnson KC. Physical activity and the risk of lung cancer in Canada. Am J Epidemiol. 2003;158(6):564–75. https://doi.org/10.1093/aje/kwg186.
55. Dela Cruz CS, Tanoue LT, Matthay RA. Lung cancer: epidemiology, etiology, and prevention. Clin Chest Med. 2011;32(4):605–44. https://doi.org/10.1016/j.ccm.2011.09.001.
56. Borch KB, Braaten T, Lund E, Weiderpass E. Physical activity and mortality among Norwegian women—the Norwegian Women and Cancer Study. Clin Epidemiol. 2011;3(1):229–35. https://doi.org/10.1186/1479-5868-9-2.
57. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. Stat Med. 2011. https://doi.org/10.1002/sim.4067.
58. Busund M, Bugge NS, Braaten T, Waaseth M, Rylander C, Lund E. Progestin-only and combined oral contraceptives and receptor-defined premenopausal breast cancer risk: the Norwegian Women and Cancer Study. Int J Cancer. 2018;142(11):2293–302. https://doi.org/10.1002/ijc.31266.
59. Kvaavik E, Lund I, Nygård M, Hansen BT. Lifestyle correlates of female snus use and smoking: a large population-based survey of women in Norway. Nicotine Tob Res. 2016;18(4):431–6. https://doi.org/10.1093/ntr/ntv126.
60. Lagerros YT, Mucci LA, Bellocco R, Olof N, Bälter O, Bälter KA. Validity and reliability of self-reported total energy expenditure using a novel instrument. Eur J Epidemiol. 2006;21(3):227–36.
61. Prince S, Adamo K, Hamel M, Hardt J, Gorber S, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. Int J Behav Nutr Phys Act. 2008;5(1):56.